

# Artificial Intelligence and People with Disabilities: a Reflection on Human–AI Partnerships



Jason J. G. White

## 1 Introduction

Having a disability gives rise to specific practical needs, related for example to mobility and transport, communication, learning, or access to information. It is an essential insight of the social model of disability [46] that the impairment of a person's capacity to function in a certain respect only becomes problematic in conjunction with specific physical or social environments. It is the combination of an impairment—for example, a sensory, physical, or cognitive limitation of the individual—with the demands of an environment that raises barriers to autonomy and social participation. Thus, by way of illustration, a physical disability that necessitates use of a wheelchair only creates difficulties if the built environment is not conducive to this mode of travel, for instance by requiring the use of stairs. Communication only poses inherent difficulties to a person who is deaf, for example, under circumstances in which the auditory mode is offered as the exclusive channel. Likewise, dyslexia only poses challenges in so far as reading is required to complete a task and suitable supports or alternatives are not available. Correspondingly, written text is problematic to a person who is blind only in the absence of a familiar, non-visual representation, such as braille or speech. Color blindness introduces challenges in so far as distinctions of color alone are used to convey information. This analysis, and the social model more generally have been criticized for offering an incomplete and inadequate conception of disability [45]. Nevertheless, its shifting of attention away from impairment as a problem of the individual that ought to be cured or alleviated as proposed by the medical model of disability, and toward the social determinants of inclusion and exclusion, is both historically and conceptually fundamental.

For this reason, much policy and advocacy in recent decades have focused on overcoming barriers to full participation in society by people with disabilities that

---

J. J. G. White (✉)  
Educational Testing Service, Princeton, NJ, USA  
e-mail: [jjwhite@ets.org](mailto:jjwhite@ets.org)

are the product of inadequacies in the design and construction of physical, social, and digital environments. The installation of ramps in buildings as alternatives to stairs and the use of braille and large print signage are among the most prominent accessibility features now increasingly found in public spaces.

Consistently with the insights derived from the social model,<sup>1</sup> article 1 of the *United Nations Convention on the Rights of Persons with Disabilities (CRPD)* [55] asserts that

[p]ersons with disabilities include those who have long-term physical, mental, intellectual, or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others.

As this characterization indicates, people with disabilities are highly diverse—an observation that will be of crucial importance in later discussion. According to the World Health Organization [59], more than one billion people, comprising approximately 15% of the global population, live with disability, and a continued increase is expected to result from the effects of aging as well as changes in the incidence of chronic health issues. Although the total population subject to disability is large, its great diversity with regard to the nature of impairments, capabilities, resources, experiences, and life circumstances undermines the reliability of naive generalizations or simplifying assumptions. Two people who appear to have a similar kind and degree of impairment may nevertheless differ greatly in their capabilities, needs, and experiences of disability. This heterogeneity is attributable to variation in the social conditions which have become the focus of critical attention by the disability rights movement and in disability studies scholarship. Indeed, the variability and complexity of the interactions between a person's impairment, development, and social conditions lie at the core of arguments against the received view that having a disability is in general bad for well-being [10].<sup>2</sup> Thus, for example, differences in educational opportunities can exercise a profound, long-term influence over the capacity of individuals not only to participate meaningfully in society, but also to address the practical needs that emerge from disability itself in such everyday contexts as employment, family life, and community activities. The quality of one's educational prospects depends significantly on socially determined conditions, including for example the availability of appropriate support, and the incidence of discriminatory treatment. The same is true in other domains of life activity.

Artificial intelligence has long served a valuable function in enhancing access for people with disabilities. Text-to-speech technology has been used as an alternative means of communication by those with speech-related disabilities. Speech recognition can function as an alternative to keyboard or pointer input, thus allowing those with certain physical disabilities to interact with software independently. From the 1970s onward, [30] people who are blind have benefited from the combination of optical character recognition with text to speech, enabling printed text to be read

---

<sup>1</sup> According to Degener [14], the CRPD acknowledges but then extends considerably beyond the conception of the human rights of people with disabilities recognized by the social model.

<sup>2</sup> The authors instead regard most forms of impairment as neutral traits that do not in themselves negatively affect quality of life.

aloud automatically. Each of these examples is an application that arguably engages capabilities that have traditionally required human intelligence. Consequently, they all amount to applications of AI, and indeed historically have constituted difficult computational problems. Each of these applications also stands to benefit greatly from recent advances in machine learning, notably deep neural networks.<sup>3</sup> It is reasonable to anticipate continued improvements in the accuracy of speech recognition and optical character recognition, as well as enhancements in the quality of text-to-speech synthesis. To this extent, the ongoing development of AI, including its shift toward machine learning, is likely further to improve the capabilities of long established applications valuable to people with disabilities.

These advances also open possibilities that have not been feasible with previous technologies. The World Institute on Disability [60] envisions automatic captioning for people who are deaf, autonomous vehicles for individuals who are unable to drive, image and facial recognition for those who are blind, language generation to support comprehension by people with cognitive disabilities, and technologies that support people with disabilities in pursuing and retaining employment. Desirable though these applications are in improving accessibility, each of them raises a variety of design challenges and ethical issues. Together, these and associated examples are taken up in Sect. 2 as suitable starting points for briefly considering some of the questions that emerge in the use of AI in overcoming barriers to access and participation. These areas of potential application focus on improving well-being by addressing challenges specifically arising from the needs and circumstances of people with disabilities.

AI presents a risk to people with disabilities in addition to opportunities. Although concerns about its potential role in discrimination have received sustained public and scholarly attention in recent years, the disability-related dimension of the problem is only beginning to be explored. Nevertheless, central issues have already been identified, and there is ample scope for further research. With this in mind, Sect. 3 offers an exploration of the problem and of some potential strategies for addressing it, emphasizing the role of the social and policy contexts. The scope of the discussion is here broadened to applications of AI in general, by which people with disabilities are affected, for example as users of a system or as individuals who are subject to decision-making processes that AI at least partly automates. Failing to address challenges of bias, discrimination, or exploitation of personal information can directly and negatively affect human well-being. There is thus a welfare-related dimension to the moral argument for establishing policies and for creating human-AI partnerships that address the potential of AI technologies to contribute to injustice. The discussion in this chapter also suggests that the concept of partnership should be understood in the current context as having social as well as technical aspects, as embracing the normative arrangements and practices in which the technology operates in addition to more narrowly conceived elements of its design and implementation.

---

<sup>3</sup> For an overview of these technical developments, see LeCun, Bengio and Hinton [31].

## 2 Use of AI to Enhance Accessibility and Inclusion

In this section, the potential of AI to contribute to the solution of practical problems arising in the lives of people with disabilities is explored. Ethical issues are raised that motivate consideration of the proper role of AI systems in given social contexts which affect design and development decisions. The general conclusions are informed by discussion of cases based on the domains of application noted by the World Institute on Disability [60]. Each of these examples is considered in turn (Sect. 2.1) to illuminate issues that it introduces. This analysis of the examples then leads in Sect. 2.2 to a reflection on the important contributions of design processes and policy incentives in influencing the fit between AI applications and the needs and values of users with disabilities.

### 2.1 *Identification of Ethical and Design Issues Through Analysis of Examples*

The examples considered here, which elaborate those put forward by the World Institute on Disability [60], serve multiple purposes. First, they are illustrative of applications of AI that have evident potential to solve practical problems which arise for people due to living with a disability. Second, considerations are introduced that demonstrate the contingency of these benefits on appropriate decisions in the design and deployment of the technologies. Such issues are the focus of the discussion in this section. Third, the examples motivate a more general treatment in Sect. 2.2 of approaches which can be taken to developing applications of AI that are genuinely and effectively responsive to the needs of people with disabilities.

#### 2.1.1 **Speech, Sound, and Image Recognition**

Captions are a well-established means of providing access for people who are deaf or hard of hearing to information conveyed in the auditory content of online video and broadcast media. Although captions have traditionally been created manually and synchronized appropriately with the video content, speech recognition and natural language processing enable this work to be increasingly automated, as the example by the World Institute on Disability [60] acknowledges. Automatic caption generation is only useful in so far as the speech recognition system is sufficiently accurate, and the degree of accuracy that ought reasonably to be required varies according to context. Thus, speech recognition could be applied as the first step of adding captions to a video as part of the production process. In this scenario, a skilled human operator is responsible for editing the captions to correct speech recognition errors. The speech recognition system need only be sufficiently accurate that it is more efficient to correct its output than to write the captions manually. However, if

the caption generation tool is used directly by a person who is deaf, for example in a live meeting, there may be inadequate opportunity to correct errors; and, plainly, the original audio is entirely inaccessible to the user. In such a situation, it is inequitable to impose the burden of dealing with the consequences of recognition errors solely on the person with a disability, except, perhaps, under conditions in which a high level of accuracy equivalent to that of human captioning can be assured. A better alternative would be to design the captioning application, or the context of its use, to facilitate manual, corrective intervention, for example by the producer of the communication or by a third party employed for the purpose, or to insist that captions be written manually by a skilled service provider. On the other hand, there may be circumstances in which a desire for privacy is best served by placing the speech recognizer under the control of the person with a disability. Whether this is the case additionally depends on privacy-related aspects of the system's design, as there is potential for disclosure of information both from the user who is deaf, and from other parties to communication. The user interface of the system could also provide notification of probable recognition errors, a capability that at least provides the user with means of judging its reliability and suitability for a given purpose in a specific context. Preliminary research investigating the styling of caption text to indicate the confidence level of the speech recognizer in the accuracy of each word suggests that this practice may be distracting to users who are deaf or hard of hearing, particularly the uninitiated [5]. It may also increase cognitive load, as the user is implicitly encouraged to interpret the confidence indicators in reading the content of the captions. There is also disagreement over whether captions should be given as full transcripts of the dialogue presented in the audio track or instead edited in the hope of reducing the required reading rate and facilitating comprehension by simplifying the vocabulary and syntax [51]. Text simplification combined with speech recognition would further increase the opportunities for error, while considerably complicating the design and development of an AI system. A decision to use automated captioning would thus tend to favor the use of full transcripts of the dialogue rather than the creation of simplified captions. In general, the appropriate course of action in deciding whether to offer automated captioning, and how the software should be designed, very much depends not just on the capabilities of the AI, but also on the social context of its proposed deployment.

The use of AI to recognize non-speech sounds and to report them via visual or tactile cues could also benefit people who are deaf or hard of hearing. However, it would evidently be inappropriate to rely on such an application in the circumstances of an emergency, for example as a substitute for providing an accessible alarm system in a building. Designing a general sound recognition tool also inherently involves making decisions about what sounds should be recognized, and what information about them should be presented to the user [18].

Similar observations apply to the use of object, face, and scene recognition by people who are blind. The question arises of under what circumstances it is appropriate to expect the person with a disability to use the technology without any effective opportunity for correction of errors, and in what conditions alternative solutions meeting the need for accessibility should be put in place. There is a risk that, unless

image recognition systems become as reliable as human observers under a wide variety of conditions, their availability will be seen as an opportunity to reduce labor costs and to impose the responsibility for addressing the effects of their limitations directly on the ultimate beneficiaries. There are also circumstances in which deciding not to use image recognition is preferable to running the risk of being misled by it. Educating users in the limitations of the technology is clearly indispensable to informing decisions about its appropriate application, whether those users are people with disabilities themselves, or other parties responsible for ensuring equitable access and inclusion, such as educators or employers. For example, misidentification of people and objects could be at least embarrassing and at worst result in ill-advised decision making—as when a stranger at the door is mistaken for a friend [18]. As with speech recognition, questions of privacy emerge, not only for the person who has a disability, but potentially for others whose information is placed at risk of inappropriate disclosure. Analogously to the case of captions, there are applications of image recognition that do not raise all of the foregoing issues. For example, it could serve as a component in an authoring tool for the development of Web sites, in which the automatically generated descriptions are supposed to be manually reviewed and corrected.<sup>4</sup> Again, however, proper application of the AI technology depends on users' knowledge of its limitations, which can be reinforced through features of the application's user interface, for example by prompting document authors to verify textual descriptions produced by image recognition.

### 2.1.2 Text Simplification and User Interface Adaptation

AI-based text simplification and summary generation tools could be valuable to people with learning or cognitive disabilities that affect linguistic understanding. However, with every simplification or summarizing strategy, there is an associated risk of misinforming and misleading the recipient, or of producing information that is more rather than less difficult to comprehend. A twofold question arises: first, how best to control this risk in the design and use of such systems, and, second, in what contexts it is appropriate to deploy these language processing technologies. As in the previous examples, AI could here be used to empower people with disabilities and to promote individual autonomy, but it could also be relied on in circumstances to which it is unsuited. The task of system designers is further complicated by the observation that the objective should be not primarily to simplify natural language as such, but rather to ensure the simplicity of the tasks that users are expected to perform with the information to be provided. As Lewis [32] has argued, simplicity should be understood as a relation between the cognitive demands of using a system, and the cognitive capabilities of the user. Thus, a system that is simple for one individual may not be so for another whose cognitive abilities are relevantly different, as Lewis [32] illustrates by analyzing the trade-offs between breadth and depth of control

---

<sup>4</sup> For further discussion of issues raised by sound and image recognition systems designed for use by people with disabilities, including some of the concerns introduced here, see Findlater et al. [18].

presentation in graphical user interface design. A deeper hierarchy, for example in a menu structure, is simpler for users who are more easily distracted, for instance, but more complex for those who have difficulty holding attention throughout a long sequence of actions. The proper role of language simplification in designing tasks that are more cognitively tractable is thus likely to be highly dependent not only on the tasks themselves, but also on the user's capabilities and on the surrounding social context.

The diverse and sometimes conflicting needs of people with disabilities regarding what constitutes an accessible user interface, as illustrated by the preceding example, have motivated efforts to develop systems that can be configured appropriately according to each individual's personal needs and preferences. Significantly, the Global Public Inclusive Infrastructure (GPII) project [57] has created software that maintains a profile of the user's needs and preferences, on the basis of which any supported system that the individual with a disability wishes to access can be automatically configured to satisfy accessibility requirements. This is achieved by matching the profile with an appropriate set of configuration choices at the operating system, assistive technology, and application levels, and then setting those parameters accordingly. The benefits of AI can be seen in the 'matchmaking' process by which the user's profile, possibly also taking into account environmental conditions such as ambient light and noise characteristics, is used to infer a suitable system configuration [27]. Two approaches to matchmaking have been developed, the first of which is a system of rules based on knowledge representation [34], whereas the second employs statistical techniques to derive a configuration by analyzing the profiles and settings of other users who have similar needs. These rule-based and statistical techniques are not mutually exclusive, and may therefore be implemented in a complementary, hybrid solution [27]. So long as privacy is preserved, this application of AI has the potential greatly to simplify and to facilitate the adaptation of user interfaces to individual access needs.

### 2.1.3 Autonomous Vehicles

As the World Institute on Disability [60] recognizes, autonomous vehicles must be universally designed if they are to satisfy the needs of users with disabilities. The features needed for a vehicle to be accessible depend on the nature and extent of its autonomy. Notably, if human participation is required in aspects of driving, as is true of all but the most fully autonomous of systems, controls and sensory feedback arrangements need to be developed which can be used effectively by people with a wide variety of abilities. While a highly accessible autonomous vehicle does not appear to have yet been created, some aspects of the problem have been the subject of preliminary research. For example, the design of tactile and auditory interfaces to enable driving decisions to be made by a person who is blind or vision-impaired has been explored [9].

The prospect of a fully autonomous vehicle occupied by a person with a disability who cannot intervene in driving to override the decisions of an AI system, raises legal

and moral concerns. Clearly, questions of safety are of paramount importance, both for the vehicle's occupant and for other road users, as are issues of ethical and legal responsibility in the event of accidents. As noted by Bradshaw-Martin and Easton [8], the use of autonomous vehicles by people with disabilities who are unable to take direct control departs from the long-standing legal assumption that a human being is responsible for driving decisions at all times. Bradshaw-Martin and Easton [8] suggest that such cases should be considered acceptable only if the operation of 'empty' autonomous vehicles (i.e., those without human occupants) on public roads is also acceptable. An alternative approach would be to enable the functioning of the autonomous vehicle to be overseen by a remote human observer who is able to assume manual control of the driving in potentially dangerous situations [8]. Although this solution introduces challenges of privacy and security, it could overcome risks to safety without greatly diminishing the independence of the person with a disability. If fully autonomous vehicles ultimately become commonplace, the skill and readiness of their human occupants to intervene can be expected to decline. Further, manual intervention is likely to be most difficult and risky in situations that pose the most danger. The problems of safety and accessibility are especially complex under conditions of mixed traffic, in which some vehicles are driven by humans and others are under the control of AI systems. In these circumstances, anticipating the actions of other 'drivers,' some of which are AI agents, can become difficult—and possibly more so for a person with a disability who is interacting with a vehicle via an assistive technology.<sup>5</sup> Fulfilling the promise of autonomous vehicles for people with disabilities thus necessitates the development of novel user interfaces, as well as a combination of technological and legal measures that can fairly allocate the risk of accidents, while reducing it to an acceptable level.

### 2.1.4 Employment and Education

The potential applications of AI that could enhance employment of people with disabilities are diverse, and capable of operating at all stages of the process from education and training to improving accessibility in the workplace.<sup>6</sup> Whether education is offered by educational institutions or directly in the work environment, there is the possibility of using AI to improve its efficacy. Intelligent tutoring systems, for example, are AI-based applications that can adapt the delivery of educational content to the learning needs of the individual. AI has also been introduced into the recruitment of employees, raising questions about the possibility of bias against candidates with disabilities. Employees may also take advantage of AI systems, including technologies supporting accessibility as considered elsewhere in this chapter, in performing their work. There are nevertheless issues of ethics and privacy to be taken into account in deciding what the capabilities of these applications should be, and in arriving at appropriate design decisions.

---

<sup>5</sup> Solutions to the general problem of mixed traffic are developed in Nyholm and Smids [38].

<sup>6</sup> See generally Employer Assistance and Resource Network on Disability Inclusion [16].



An ‘intelligent’ educational application could, in principle, adapt the presentation of material and its evaluations of the learner’s responses according to needs arising from a disability. For example, it could offer additional explanations of geometric or spatial concepts to a student who is blind and whose knowledge of the relevant spatial relationships is found to be in need of consolidation. The possibility of individualizing the delivery of education based, in part, on a person’s disability has genuine potential to improve learning, leading ultimately, in the present context, to greater success in a career. However, it also has the potential to perpetuate misconceptions about what people with disabilities can do and to entrench stereotypes. For example, whether a student who is blind would benefit from additional support in performing tasks requiring geometric knowledge ought not to be inferred from the disability category, but should instead be ascertained with respect to each person individually. Similarly, whether sign language interpretation should be provided for multimedia content (for instance, in a tutoring system) depends on the individual’s knowledge of and preference for a sign language—factors that are not captured by the classification of the individual as a student who is deaf. There is thus a risk of drawing inappropriate generalizations from a disability classification, instead of attending to the specific needs of each learner. Adaptive AI systems developed as educational technologies could exacerbate this problem, unless suitable design choices are taken.

Further cause for concern emerges from the possibility of building educational systems that use AI-related techniques in an attempt to determine whether a person has a disability. For example, an arithmetic tutoring application might be equipped with the ability to flag a student as possibly having dyscalculia—a learning disability. The negative personal and social consequences that could result from a misclassification are considerable, including stigmatization and inappropriate educational interventions.<sup>7</sup> The same could also occur in the event of a correct classification, particularly in the absence of knowledgeable and skilled educators who understand the nature of the disability and the needs of the student. Depending on the design of the system and the conditions of its use, individual privacy rights could also be infringed in this scenario. In general, the design of AI systems to detect a disability—particularly a disability of which the individual may be unaware—is fraught with ethical difficulties, while also giving rise to legal issues. For example, the European Union’s General Data Protection Regulation (GDPR) ([17], article 5(1) (b)) constrains the processing of personal information for purposes that are incompatible with the ‘specified, explicit and legitimate’ purposes for which the data are collected, and consent to which is among the permissible bases of authorization ([17], article 6(1) (a)).<sup>8</sup> In addition, the creation and disclosure of health-related information is a sensitive matter that is appropriately subject to legal safeguards which vary by jurisdiction.

---

<sup>7</sup> The risks of using AI as a tool of medical diagnosis in relation to people with disabilities is discussed in Trewin et al. [54].

<sup>8</sup> The limitation of data processing to specified, explicitly stated purposes is an aspect of European data protection law that raises difficulties for machine learning-based AI applications generally. See Marsch [37] for treatment of the relevant human rights obligations.

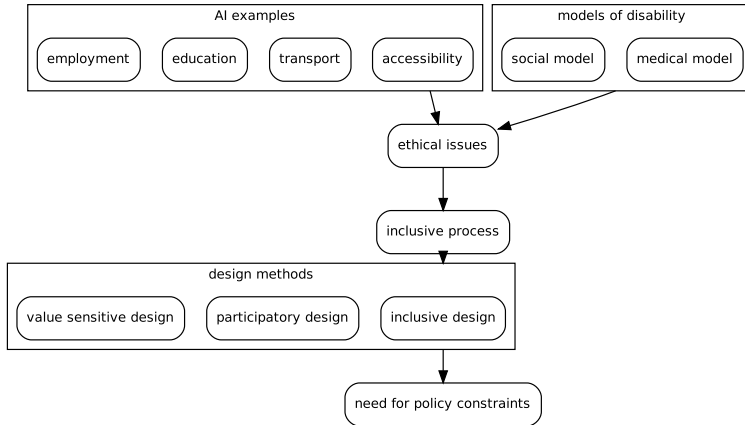
A further case which well illustrates the importance of preserving privacy is that of a hypothetical recommender system designed to match prospective employees with available employment opportunities. Under readily foreseeable conditions, voluntary disclosure of an individual's disability status to the system could be beneficial. For example, in some countries, there are policies in place which establish quotas to improve the employment rate of people with disabilities, who are expected to comprise a specified proportion of each employer's workforce [22]. A recommender system could take an individual's disability status into account, together with qualifications and experience, to suggest opportunities offered by employers who are likely to have unfilled quotas. Of course, this would require informed consent, as some people with disabilities may object to having their disability status operate as a factor in an employment decision. On the negative side, however, any disclosure of disability status to employers by such an AI system not only raises privacy issues, but also creates a risk of discrimination in the selection process. Such a case is thus illustrative of the value of statutory protection of privacy, and of requiring consent for the purposes for which data are collected. Controlling the disclosure of information—in this case, about a person's disability—thereby limits the opportunities for its misuse.

## 2.2 *Observations*

Drawing on a recent paper [47] that connects technological choices with social assumptions revealed by the contrasting models of disability introduced in Sect. 1, it is argued in Sect. 2.2.1 that a collaborative and participatory approach is necessary, which engages people with disabilities directly in the development of AI systems designed to meet their needs. This position is supported by further comments on examples considered in Sect. 2.1. Section 2.2.2 takes up the suggestion, advanced in a recent contribution to the literature on AI and people with disabilities [54], that theoretical and methodological traditions in design thinking have much to contribute to an elaboration of what constitutes appropriate participation. This scholarship also offers a framework—value-sensitive design—in which to address the moral questions raised by AI-related projects. In Sect. 2.2.3, it is maintained that such design-based approaches, though valuable, should also be complemented by supporting norms and incentives grounded in policy. The principal conceptual relations developed in this section are depicted in Fig. 1.

### 2.2.1 **The Need for an Inclusive Collaboration**

Each of the examples in Sect. 2.1 is a good illustration of the potential benefits that AI can bring to people with disabilities. It can serve a positive function by helping to overcome problems of access to information, communication, education, employment, and transport. Nevertheless, as has been shown in each case, there emerge important questions to be considered in deciding what AI applications should



**Fig. 1** Examples of AI applications intended to benefit people with disabilities (in accessibility, transport, education and employment), together with the medical and social models of disability, raise ethical issues. These issues give rise to the need for an inclusive development process, supported by design methods such as inclusive design, participatory design, and value-sensitive design. Consideration of these design methods illuminates the need for policy constraints

be built, and in arriving at appropriate design decisions. As has been emphasized in the discussion, these issues are concerned not only with the AI system as a technological artifact, but also with the wider social environment in which it is likely to be deployed and used. Choices about design and implementation should thus be made from a perspective that is informed by knowledge of how a proposed system would actually function in the specific life circumstances of people with disabilities.

In deciding what problems AI should be used to solve, and what constitute adequate solutions, there is a risk of introducing prejudiced assumptions about the lives of people with disabilities. Shew [47] warns of this danger, noting examples of technologies that perpetuate problematic assumptions about disability, particularly the notion central to the medical model that the shortcoming essentially resides in the bodily limitations of the individual, which must be ‘treated’ or minimized, rather than in the physical and social context. Narratives about the benefits of technology for people with disabilities are connected with notions of independence, which, Shew points out, downplay the extent to which people in general are interdependent in a multiplicity of respects. Thus, the promise of autonomous vehicles as enhancing independence for people with disabilities reinforces these narratives, situating the problem in the individual’s inability to drive, rather than in a social responsibility to provide effective and accessible means of transport [47].<sup>9</sup> A more careful analysis would recognize that although autonomous vehicles would increase independence in some respects, they would also create a less obvious dependence on the developers

<sup>9</sup> Shew further develops the point in a brief discussion of additional examples, including the rationale for using companion robots, which may serve the interests of human care givers more than those of the person with a disability whose needs are to be met.

and maintainers of the technology—designers, implementers, trainers, and service personnel among them. Thus, the central question is concerned not with a greater or lesser degree of independence as such, but instead with the type of interdependence that is desired. An autonomous vehicle offers freedom to choose when and where to travel, without having to coordinate with other people (namely human drivers or public transport operators). However, it also requires the person with a disability to entrust his or her safety to the creators and maintainers of a complex AI system.

The discussion of speech, sound, and image recognition in Sect. 2.1.1 directs attention to issues of societal responsibility for making information and communication accessible, as well as associated questions of privacy and confidentiality. Clearly, imposing responsibility for using AI to solve problems of information access principally on the individual who is faced with accessibility barriers is consistent with an individualistic, medicalized concept of disability. Also, as has been noted, it shifts the burden of the technology's shortcomings onto the person who is least able to overcome them. This is not to suggest, however, that people with disabilities should be deprived of opportunities to gain the full benefit of such technologies and to use them independently. Rather, the point is to acknowledge the need for informed decision making about appropriate application, both at the individual level and in policy decisions regarding the overcoming of obstacles to access and inclusion.

Shew emphasizes the importance of ensuring the agency and autonomy of people with disabilities, and of fully recognizing their expertise in their own needs, lives, and experiences. This can be achieved, in part, by bringing the knowledge and experience of people with disabilities directly into the process of making decisions about the development and use of AI applications. User participation in design is thus of crucial value, as is educating decision-makers and developers about disability.

The concept of human–AI partnership can be extended to acknowledge the importance of social practices and institutions in shaping the creation and application of AI-based technologies. What is proposed here is that technological decision making ought to be sensitive to this social context, and, ideally, not isolated from choices about the social practices in which technologies are developed, maintained, and used. Questioning implicit presuppositions, as well as gaining a greater understanding of the lives, desires, and needs of people with disabilities are essential aspects of this approach. The slogan put forward by the disability rights movement, ‘nothing about us without us’, aptly conveys the importance of involving people with disabilities directly in making decisions that affect them, including choices about appropriate uses of AI technologies [58].

The considerations advanced so far build a case for engaging people with disabilities directly in problem identification and definition, as well as in determining what constitute appropriately designed solutions capable of meeting their practical needs effectively. To be clear, it is not argued here that only people with disabilities are appropriately qualified to conceive and to plan suitable AI-based solutions. Rather, the claim is that the design and development of these systems should be carried out in close collaboration with people who have disabilities and should preferably be undertaken by developers who are personally interacting with disability-related communities in non-trivial ways. Through a genuine and mutual understanding of

the problems to be solved and of the design possibilities, informed decisions can be made which lead to AI projects that are successful in enhancing equity and life opportunities. The risks of perpetuating prejudiced assumptions, and of devising systems which attempt to solve the wrong problems, can thus be minimized by establishing an inclusive collaboration in which prospective users and beneficiaries play a crucial role. This collaboration should persist throughout a project, from its inception through to the evaluation and refinement of the delivered AI system.

### 2.2.2 The Role of Users in Application Design

As discussed by Trewin et al. [54] in connection with the development of AI systems that avoid algorithmic bias against people with disabilities, there are traditions of research and practice in which the users of a technology play a central role in the design process. Among these traditions, the authors emphasize the contributions of ‘inclusive design,’ ‘participatory design,’ and ‘value-sensitive design’ in particular. Such approaches are clearly germane not only to the design of AI applications generally, but also to the development of applications which, in their over-all purpose or via the inclusion of certain assistive technologies and accessibility-related features, aim to satisfy needs specific to certain users who have disabilities. In these cases, the potential users with disabilities who are intended to benefit from a proposed project (an automated captioning application, for example) can be identified, and efforts can then be made to engage representative individuals. Drawing on design methods that privilege the user’s perspective in decision making or which make explicit the value judgments inherent in technical choices has the potential to improve the quality and suitability of the resulting AI systems. Nevertheless, the limitations of these methods should also be kept in mind in relation to the project at hand.

In participatory design, for example, the users are genuine partners in decision making. Participatory design approaches originated in the Scandinavian industrial democracy movement, and were substantially motivated by resistance to Taylorism in the workplace.<sup>10</sup> There thus arose a tradition that recognizes the value of the tacit knowledge possessed by users (in the original context, industrial workers) in performing tasks and solving problems. Instead of seeking to analyze, formally describe and optimize this tacit, practical knowledge, participatory design develops and builds upon it in ways that are meant to empower the users and to preserve their autonomy. Moreover, participatory design methods are intended to be applied to the entire work process, not merely to the creation of technological artifacts. The emphasis placed on preserving and enhancing users’ tacit knowledge gives rise to a tendency toward solutions that retain instead of radically reconfiguring established practices—a favoring of evolutionary over revolutionary change.<sup>11</sup> These aspects of the approach have

---

<sup>10</sup> For an overview of the history and the guiding ideas, see Ehn [15]. A more recent introduction to participatory design appears in Spinuzzi [49].

<sup>11</sup> This consequence of the value placed on preexisting tacit knowledge is acknowledged as a limitation of participatory design in Spinuzzi [49].

the potential to contribute to improving the design of AI systems intended for use by people with disabilities. However, they may also discourage more fundamentally innovative, long-term projects.

Development of an AI system to facilitate indoor navigation by people who are blind, for example, would begin with an understanding of existing, formal and informal practices of orientation and mobility. It would tend to favor extending established approaches, such as those centered on the white cane and the guide dog, instead of proposing the development of robots as substitutes. It would also seek to preserve existing mobility skills. This could be achieved, for instance, by further developing orientation applications suitable for use with mobile phones or wearable devices that could give information and directions to the user, especially in unfamiliar, indoor settings, and ideally without requiring the installation of specialized infrastructure such as radio beacons that enable precise locations to be identified. On the other hand, developing robots capable of navigating and of guiding their users in a wide variety of environments is arguably a valuable, long-term objective, notwithstanding the practical limitations of current prototypes. The state-of-the-art prototype described by Guerreiro et al. [23] is largely limited to flat, indoor environments in virtue of technical constraints, including weight and battery capacity. If these constraints could be overcome and the capabilities of the AI system responsible for navigation improved, the technology could manifestly be advantageous to users, at the risk of their placing too much reliance on a robotic guide to the detriment of existing skills. The shortfall in the user's orientation and mobility skills would then become problematic whenever robotic assistance was unreliable or unavailable. The principal benefits of robots—effective navigation in unfamiliar settings, and avoidance of hazards or obstacles that conventional mobility aids would miss—could be obtained by developing solutions that extend rather than supplant current tools and strategies. The robots, however, may in the long-term offer usability advantages that would not be achieved via more evolutionary approaches.

Respect for users' tacit knowledge in the design of new technological solutions is thus autonomy-preserving, enhancing the individual's control over the manner in which tasks are performed and capacity for decision making. It operates also as a constraint on more radical forms of innovation. Participatory design offers the advantage of making the practical knowledge and skills possessed by users conspicuous, and therefore of raising questions about the role it should continue to occupy in the application of new, AI-based systems. The fostering of user involvement in design decisions also provides practical means of negotiating these issues, among others, in arriving at appropriate technical solutions.

The design problems that have here been discussed all introduce questions of value, broadly construed. Value-sensitive design seeks to bring investigation of the values implicated by technological choices directly into the development process.<sup>12</sup> Importantly, value-sensitive design attends to the salient moral considerations that

---

<sup>12</sup> The approach is articulated and illustrated in Friedman, Kahn and Borning [20]. For a recent treatment of the underlying concepts and design methods, see Friedman and Hendry [19] (Chaps. 2 and 3).

may otherwise be overlooked or disregarded in technological projects. It accordingly has potential to serve as a useful tradition to draw upon in building AI systems for people with disabilities. Through ‘conceptual,’ ‘empirical,’ and ‘technical’ investigations, value-sensitive designers identify and engage direct and indirect stakeholders—people whose interests are affected by a project. The values of the stakeholders and of the designers themselves with respect to the design problem are investigated and incorporated into technical decisions. Of particular significance in relation to the role of AI in enhancing the lives of people with disabilities is the concept of ‘value tensions’ [19] (Chap. 2) among and even within stakeholders. Recognition of these value tensions can lead to creative solutions that reconcile apparently conflicting priorities. For instance, [19] (Chap. 2) an energy-efficient design may be agreed upon both by stakeholders who prioritize cost minimization and by those who favor environmental sustainability, without requiring resolution of their underlying moral disagreement. Of course, this is only possible if the proposal contains costs while also reducing energy consumption from sources that are ecologically harmful.

Opportunities for creatively overcoming value tensions can and indeed should be sought in cases such as those discussed in this chapter. As an illustration, the choice between providing captions that give a full transcript of the dialogue in a video, and providing captions as simplified summaries of the dialogue, can be understood as a value tension that may arise not only between users but even within a single individual. As is evident from the discussion in Szarkowska et al. [51], unsimplified captions offer the user full access to the dialogue, without interposing another person’s interpretation of it, whereas simplified captions can improve readability and comprehension, which are also valuable to and valued by users, at the cost of precluding equal, unmediated access to the spoken content. If sufficiently reliable speech recognition and summary generation technologies were available, this value tension could be overcome simply by generating both types of caption. The user could then choose which type of caption to read in each particular situation. Though attractive, this solution also exacerbates a second value tension—that between the quality and availability of captions, on the one side, and the desirability of containing production costs by reducing the human labor associated with editing and verifying captions, on the other. Much of the appeal of automating caption creation derives from the desire for improved cost efficiency. Generating two sets of captions for video content obviously runs contrary to this objective. The resulting tension can be resolved by a technological solution if the automatic speech recognition and text simplification algorithms are sufficiently accurate to maintain labor costs that are acceptable to the producers. This places a heavy demand on AI researchers and software developers. In the absence of a technical solution, the value tension should instead be regarded as a value conflict. The wider question, then, is how to address such value conflicts if, as in the current example, the human rights of people with disabilities are at stake. In this case, the right of access to information and communication [55] (article 9) is implicated.

### 2.2.3 Commentary

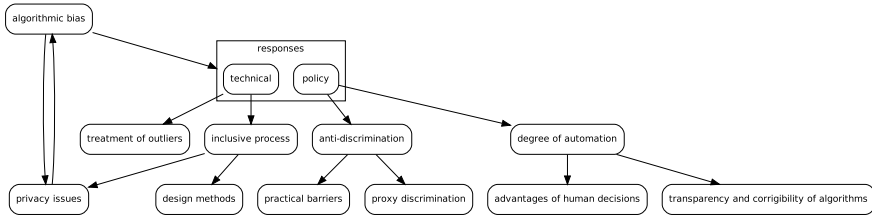
Engaging users appropriately in the design process has genuine potential to encourage the development of technologies that are well aligned with the needs and values of people with disabilities. However, treating the resolution of value tensions largely as a design question for negotiation among designers and stakeholders in individual technical projects would risk the formation of compromises that undermine the rights and interests of those whom it is supposed to benefit. For this reason, policies, ultimately established by governments, have a necessary and important role. Indeed, the presence of deeply entrenched practices of social subordination operating against people with disabilities, manifested in technological choices as Shew has described, justifies skepticism toward the ability of designers and stakeholders to resolve value conflicts appropriately in the absence of incentives created by policy. In AI, as elsewhere, one can plausibly argue that regulation and oversight are indispensable elements of upholding a moral commitment to social equality for people with disabilities. Trewin et al. [54] do not suggest otherwise. Nevertheless, this conclusion is important in understanding the complementary role that design methods occupy to policy considerations in relevantly shaping the future of AI development.<sup>13</sup> Furthermore, design procedures centered on the participation of users and other stakeholders are expensive, raising doubts about whether organizations responsible for building AI-based technologies will deploy them sufficiently in the absence of externally imposed incentives. True power sharing among stakeholders in the design process challenges existing structures of authority in addition to creating participation costs. These factors suggest that its widespread application to AI development will require policy-based interventions. In evaluating current policies and planning future regulatory approaches, there also arises the challenge of promoting the rights of people with disabilities while allowing suitable flexibility for stakeholders to arrive at creative solutions to technological problems, including mutually advantageous responses to value tensions.

The development of human-AI partnerships that respect the rights and satisfy the needs of people with disabilities thus requires an interplay of technical and social choices. The social aspects of these choices bring to the fore issues of morality, including questions of justice and human rights. Adequate resolution of these considerations in technological projects depends on the nature of the design process, the motivations and skills of the participants, as well as the internal and external incentives that influence decisions. There is an important role for policy in establishing appropriate incentives. Traditions emerging from design research can also be applied and refined to support meaningful participation by direct and indirect stakeholders with disabilities. Addressing value tensions can be regarded as partly a function of the design process, and in large measure as lying in the domain of overarching policies.

---

<sup>13</sup> Interestingly, Friedman and Hendry [19] (Chap. 2) regard policy as a kind of technology for the purpose of applying value sensitive design methods.





**Fig. 2** Mutually related problems of bias and privacy give rise to technical and social responses. Technical responses concern treatment of outliers and inclusive development processes. Inclusive processes engage appropriate design methods (as described earlier in the chapter), while also interacting with privacy issues. Policy responses concern antidiscrimination (practical barriers as well as normative issues of proxy discrimination) and questions concerning the appropriate degree to which decisions should be automated. The latter questions, in turn, raise issues regarding the advantages of human decision making, and the transparency and auditability of machine learning algorithms

### 3 Disability as a Site of Algorithmic Bias

Whereas the previous section examined respects in which AI systems can be designed to benefit people with disabilities by solving specific, practical problems, the following discussion addresses the more general issue of the role of AI in perpetuating, and even amplifying, discrimination against them. The disability-related aspects of the problem of algorithmic bias are introduced (Sect. 3.1) and briefly illustrated by citing examples presented in recent literature (Sect. 3.1.1). Closely related issues of privacy are discussed in Sect. 3.1.2. The difficulties associated with technical, social, and policy-related remedies are then explored (Sect. 3.2), enabling the identification of open questions pertinent to research and practice. Even more so than in the preceding section, the purpose of this commentary is to pose questions rather than to recommend solutions and to offer a conceptual approach to thinking about the problems rather than to give concrete guidance. The accumulation of multidisciplinary research and evidence derived from actual cases of AI applications in subsequent years can be expected to clarify the issues, while strengthening the guidance available to practitioners. Figure 2 presents a conceptual overview of the issues considered.

#### 3.1 Introduction to the Problem of Bias

It has long been established that software, including AI applications, can reinforce biases already present in the social context, while introducing new sources of bias of its own [21]. The growing utility and increasingly diverse applications of AI systems based on machine learning which have emerged in recent years greatly expand the potential for biases to be introduced and extend the range of possible harms that may result. As decision making becomes increasingly automated in a wide variety of domains, all the more opportunities arise for biased algorithms to contribute to

social injustice,<sup>14</sup> including discrimination against people with disabilities. Questions concerning algorithmic bias with respect to disability are necessarily part of a larger discussion of the role of AI in discrimination generally. What distinguishes disability from other social categories subject to algorithmic bias, such as gender and national or ethnic origin, consists in the nature of the diversity that disability represents. As introduced in Sect. 1, impairments can affect a broad range of human functioning, including sensory, physical, psychological, and cognitive aspects. A given individual can have one or more impairments of different kinds and degrees, which may occur at different stages of life and vary over time. These impairments can then combine with highly variable, socially mediated conditions to constitute disabilities that present practical challenges to the individual. As with other social categories of interest in connection with algorithmic bias, societal practices of subordination and exclusion can have a large role in limiting a person's well-being and the life opportunities that can effectively be pursued. As has been recognized in recent literature, [53, 54, 58], the great diversity of impairments, social circumstances, resources, and experiences among people with disabilities creates an associated diversity in the ways in which they can be subjected to biases in AI systems.

### 3.1.1 Potential for Bias

Scholars and practitioners have reported on the findings of workshops that have identified a variety of AI applications in which such biases have real potential to occur [54, 58]. These examples are not exhaustive, as the range of AI applications is large and becoming more so. Nor are all of the examples reviewed here, as the discussion which follows can be sufficiently motivated by a brief overview.<sup>15</sup> It is clear from the actual and hypothetical cases discussed in the literature that the biases in question tend to operate against people with specific circumstances and types of disability, rather than against people with disabilities as a general category. Moreover, an AI system exhibiting bias may do so to different extents and in different ways to different individuals. This is a product of the many dimensions of diversity characteristic of people with disabilities. A clear illustration of the specificity and the danger of algorithmic bias is given by Treviranus, who presented machine learning models designed to control autonomous vehicles with images of a friend who 'propels herself backwards in her wheelchair' [52, 1]. The models would have directed the vehicle to run her over at an intersection, and, worse, they reached this decision with even greater confidence after having been trained with data depicting people in wheelchairs [52].

AI systems designed to draw inferences from the behavior or appearance of a person are problematic, since people with disabilities can differ in many respects from the relatively homogeneous populations likely to be used in training. Examples

---

<sup>14</sup> An informative overview of how discrimination can occur is presented in Barocas and Selbst [3].

<sup>15</sup> The cited references should be consulted for more detailed illustration and discussion of applications in which bias against people with disabilities can reasonably be foreseen.

of these applications, and the ways in which they may adversely classify people due to a disability, have been noted in domains such as partially automated job interview assessment, and public safety systems meant to detect suspect behavior [54, 58]. The monitoring of a person's interactions with a user interface, for example in an educational measurement application such as a test or an interactive learning tool, raises similar concerns [54, 58]. More generally, machine learning algorithms have been applied to various tasks in which people are ranked or categorized to determine their eligibility for an opportunity or benefit, or their liability for a sanction. In employment, for example, AI has been applied at every stage of the process from deciding whom to select for targeted advertising of open positions, to the screening of job applications, and ultimately the monitoring of the employee's work performance [7]. In such cases, biases could occur against people with disabilities for different reasons and to varying extents, depending on details of the interactions between disability-related circumstances and the factors taken into account by the machine learning algorithm. Such algorithmic bias may then be reflected in adverse decisions with discriminatory effects [54, 58]. For instance, an individual's job application may be automatically excluded from further consideration or an employee's work performance may be automatically flagged as likely to be inadequate.

As an additional example, AI systems have the potential to be used extensively by governments to determine eligibility for welfare benefits.<sup>16</sup> This prospect raises the possibility of algorithmic biases that could disadvantage people with disabilities, especially those who are in greatest need of public support. A formula used by software to calculate individualized budgets for government-funded services needed to support the independent living of adults with developmental disabilities has become the subject of litigation in the USA on grounds of due process [56]. Although this case does not appear to be an example of an AI technology, it is indicative of the types of welfare-related decisions that could be readily carried out by machine learning applications.

### 3.1.2 Privacy and Bias

There is also a complex relationship between privacy and the problem of bias in AI applications. Inclusion of data obtained from people with disabilities is often necessary to the construction and evaluation of machine learning systems that avoid or minimize bias. However, the acquisition of information that reveals a person's disability also introduces opportunities for exploitation, or for unintended but nonetheless substantive discrimination, whether carried out by the data collector or by third parties to whom details are disclosed. As noted in Trewin et al. [54], this problem is further complicated by diversity among people with disabilities. The exclusion of obviously identifying information from data collections may not be sufficient to anonymize them. Knowledge of the person's disability, combined with other attributes, may be enough to enable the individual to be uniquely identified, for example as the only

---

<sup>16</sup> See Alston [1] for an overview of human rights-related concerns about this practice.

wheelchair user who lives in a given locality [54]. Some individuals, such as those with cognitive disabilities that preclude the requisite understanding, may be unable to give informed consent to the acquisition and use of their data. Yet, these data could be highly valuable and indeed indispensable to the development of AI applications designed to enhance the well-being and to improve opportunities in life for such populations.

An additional risk of discriminatory treatment is created by what Marks [35] refers to as ‘emergent medical data,’ namely health-related information about an individual that is inferred at a high degree of probability from diverse sources of evidence. The distinctive characteristic of emergent medical data is that none of the sources of evidence is overtly health-related. Consequently, no voluntary disclosure of medical information is involved. For example, a person’s purchasing history could be combined with indicators gleaned from online communications and interactions with social media applications to infer the nature of the individual’s disability [36]. This disability classification, whether or not it is accurate, could then serve as a ground for unjustly denying a benefit or opportunity. Marks is essentially concerned with the deliberate inferring and later exploitation of disability-related information. However, as noted in Sect. 3.2.2, this process could also take place unintentionally. Whereas intentional derivation and misuse of medical data can be regulated by privacy laws as Marks [36] discusses, the possibility that a machine learning system could detect and respond adversely to disability in a completely autonomous fashion raises additional difficulties.

### 3.1.3 Summary and Comments

Thus, it is clear that the nature of the diversity manifest among people with disabilities opens the possibility of bias in a variety of AI applications, especially those built on machine learning. Being a statistical outlier—one who is significantly different in a relevant respect from most of the population—can readily lead to misrecognition or misclassification of a person by a machine learning model. For this reason, technical measures that have been proposed to address the problem of AI bias against people with disabilities focus largely on improving the ability of machine learning systems to treat outliers appropriately [53, 54]. The distinct but related problem of maintaining adequate privacy protection for information that reveals aspects of a person’s disability also calls for technical and regulatory solutions. These solutions are necessary to support the acquisition of data enabling people with disabilities to be included in the development of machine learning systems, thereby alleviating bias and consequent discrimination. However, privacy controls can also reduce the risk of biases that result from the exploitation of emergent medical data.

## 3.2 *Responses to the Problem of Bias*

By reviewing some of the potential measures that can be taken to avoid or to mitigate bias in AI systems, it is possible to identify research problems of particular relevance in the context of disability. What follows is not therefore intended as a comprehensive survey of possible interventions, but rather as a discussion of starting points in this direction which illuminate issues worthy of further investigation.

Technical approaches to overcoming bias suitable for adoption in software development projects are briefly noted in Sect. 3.2.1. Attention is then turned in Sect. 3.2.2 to the limits of antidiscrimination law as a regulatory solution, emphasizing the challenges introduced by proxy discrimination and its relevance to decisions affecting people with disabilities. In determining which practical problems to solve by means of AI and in weighing the adequacy of proposed solutions, choices often need to be made concerning whether, how and to what extent decision making in the relevant domain of application should be automated. Issues concerning the strengths and weaknesses of human and algorithmic decision making are raised in Sect. 3.2.3 as they arise in relation to the automation of decisions involving a highly diverse population. Concluding remarks appear in Sect. 3.2.4.

### 3.2.1 **Technical Measures**

The advice for developers of AI systems put forward in Trewin et al. [54] is aligned with the typical process of building a machine learning application. Emphasis is placed on systematically identifying people with disabilities who constitute potential outliers for purposes of the application under development, and including them at all stages, beginning with the planning of the project and progressing through to testing of the delivered product. Once the application is deployed, monitoring of its outcomes and remediation of any discovered biases are recommended. Crucially, the inclusion of people with disabilities consists in both engaging them directly as part of the project and incorporating their data into the design and training of machine learning models. Attention is paid to questions of privacy, noting standardization efforts toward developing technical controls that can be implemented to enable users to specify their privacy-related preferences. As was discussed in Sect. 2.2, the authors recommend drawing on traditions of design practice in which the involvement of users and other stakeholders is accorded a central role. The guidance offered in Trewin et al. [54] serves as a valuable point of reference for anyone who is concerned with the practical challenge of designing machine learning applications which are inclusive of people with disabilities.

There are practical limits to the number and therefore the diversity of users or other stakeholders who have disabilities that can be meaningfully included in a software project. The people with disabilities who are introduced into the process should therefore be regarded as having a representative function. Their contributions of data and insight derive from their own personal circumstances and experiences

of disability. They may also be able to deploy personally or professionally acquired knowledge concerning others who have disabilities and whose backgrounds differ relevantly from their own. The knowledge possessed collectively by participants in a project, including people with disabilities themselves, can thus vary substantially, even if systematic efforts are made to include appropriate stakeholders. How well this knowledge represents the actual diversity of the population who will ultimately be subject to the AI system may be decisive in determining the extent to which biases are avoided or minimized. For example, it is entirely plausible that an autonomous vehicle development project which effectively and meaningfully engaged wheelchair users at every stage could nonetheless overlook individuals such as the friend described by Treviranus [52]. The diversity among people with disabilities thus creates a challenge of representativeness and of collective expertise in AI-related software projects, even under favorable conditions in which inclusive development practices are followed. To what extent and under what circumstances engagement of suitable stakeholders with disabilities can effectively mitigate bias in machine learning systems should hence be regarded as an open research problem. It is also a strategy that holds considerable promise. The recognition that it raises unresolved research questions by no means diminishes its practical importance.

### **3.2.2 The Role and Limits of Antidiscrimination Law**

Technical approaches can thus be taken to avoid the introduction of bias and to remediate it if it is detected in operational applications. Of course, these technical solutions are only likely to be implemented if appropriate social conditions are established, including incentives to undertake the necessary design and development work, and to do so competently. Antidiscrimination law is a major source of this incentive. However, there are also grounds for skepticism about the ability of antidiscrimination law effectively to regulate algorithmic bias against people with disabilities. The first consideration is practical: antidiscrimination laws are typically enforced only in response to proceedings brought by people who claim to have been subjected to unlawful discrimination. Bringing such a complaint requires one to engage considerable expertise and resources in challenging decisions made by or with the support of an AI system. Such advocacy may be problematic due to an individual's circumstances—for example, socioeconomic disadvantage and shortfalls in the availability of free or low-cost legal representation. Disability, including past practices of discrimination, can readily exacerbate difficulties that operate against bringing an antidiscrimination claim. In addition, individuals who, due to the nature of their disability, cannot participate directly in bringing a claim depend completely on others to assert their rights and to represent their interests. A public authority empowered to monitor potentially discriminatory AI systems, to investigate their operation, to respond to complaints, and to require adherence to legal standards of non-discrimination, could overcome the limitations of relying entirely on individual claims as an enforcement mechanism. Indeed, such a regulator—a 'neutral data arbiter'—has been proposed to address privacy-related harms associated with the

use of data analytics [12, § III]. Its role could readily be extended to questions of non-discrimination, including those associated with disability.

The second consideration is that machine learning can be biased in ways that raise difficulties for regulation by antidiscrimination law. These difficulties also create practical challenges for designers and developers of AI applications in avoiding bias.<sup>17</sup> According to Prince and Schwarcz [41], machine learning applications that combine data from a variety of sources have a propensity to lead to proxy discrimination. Proxy discrimination occurs under circumstances in which a protected characteristic such as race, gender, or disability status is actually predictive of the legitimate outcome of interest to the discriminator, in ways in which other variables are not. The authors argue that, since the goal of machine learning is to optimize predictive accuracy with respect to the target variable based on the input data,<sup>18</sup> it can be expected to ‘discover’ unobvious correlates of protected characteristics even if those characteristics and their obvious correlates are excluded from the data in an attempt to prevent bias. Thus, they suggest, [41, § I] a person’s membership in an online forum devoted to a particular genetic medical condition could lead a machine learning algorithm to recommend a higher insurance premium. In this case, membership in the forum is predictive of the target variable, namely insurance risk. There is a causal connection that runs from having the medical condition or a close relative who does, to both joining the online forum and having an elevated risk of disease. These correlations are of course far from perfect, but the point is that they are causal and sufficiently significant to be predictive. Forum membership is thus an unobvious proxy for sensitive health information that is excluded from the data supplied to the machine learning algorithm. One may further suppose that the medical condition could in turn be predictive of acquiring a disability.<sup>19</sup>

The fact that proxy discrimination is taking place may be entirely unknown to the discriminator and indeed to all the developers and users of the machine learning system [41]. Furthermore, due to the great diversity among people with disabilities and the unobvious correlates of disability-related information that may be present in data used by machine learning algorithms, the problem presented by proxy discrimination has the potential to be particularly difficult in this context. Identifying and excluding or otherwise addressing unobvious proxies for disability-related information that is genuinely predictive of the target variable stands as a technical challenge. Unlike Marks’s concern with emergent medical data (Sect. 3.1.2), which are derived and used intentionally, the probabilistic inferences that lead to proxy discrimination arise internally to and as the product of the ‘normal’ operation of machine learning systems. They may be unintended, and they may also be difficult to detect. There is

---

<sup>17</sup> Trewin et al. [54] acknowledge the practical dimension of the problem, and recommend consultation with stakeholders as part of the development process.

<sup>18</sup> The target variable is that which the machine learning model is designed to predict. It is assumed here to be in the legitimate interest of the discriminator, such as the probability that a person would be an effective employee.

<sup>19</sup> Hoffman [24] argues that anti-discrimination law should be extended to address decisions based on predictions of a person’s likelihood of developing a disability, and to require disclosure of the use of data in making such decisions.

thus an open research problem concerning the extent to which and the respects in which proxy discrimination is a particular difficulty in machine learning applications affecting people with disabilities, as well as what measures can be taken to control it. The diversity of impairments and the variable social conditions that affect the lives of people with disabilities provide a good ground for hypothesizing that proxy discrimination could here pose a substantial challenge. This challenge is further complicated by intersectional considerations that result from the multiplicity of legally protected social categories to which a single individual may belong. To extend the example, suppose not only that a person's online forum membership is a proxy for having a given medical condition associated with a disability, but also that there are linguistic indicators in her or his contributions to social media which function as proxies for belonging to a marginalized ethnic minority in the country in which he or she lives. Suppose further that, due to discrimination in the provision of early diagnosis and treatment services, the conjunction of having the medical condition and belonging to the ethnic minority is strongly predictive of adverse health effects of interest to the insurer, whereas neither circumstance is significant alone. Under such conditions, the disability is an essential factor in the proxy discrimination, but it only operates in combination with other category memberships. Apart from the technical and practical difficulties that such possibilities raise, there may also be legal obstacles, for example if the law requires one to choose which ground of discrimination to assert. In the current example, a choice may need to be made between alleging disability and racial discrimination, neither of which is well suited to the case.<sup>20</sup>

The legal difficulty which proxy discrimination creates is not confined to the empirical issue of establishing sufficient evidence of discrimination. Proxy discrimination also entails that eliminating variables which explicitly represent disability-related information as well as obvious proxies for them from the input data is inadequate to prevent bias [41]. Alleged discriminators can also seek to justify their practice by arguing that, since the proxies relied on by the machine learning model are truly predictive of the outcome of interest, and the model has been optimized for predictive success, no less discriminatory alternative is available that would be equally effective in achieving the defendant's legitimate objective. As Prince and Schwarcz [41](§ IV.A.2) argue in relation to disparate impact doctrine in the USA, this reasoning, if it is found to hold according to the facts of a particular case, can serve as an adequate defense against a claim of unlawful discrimination. Clearly, whether this is so depends on the details of the antidiscrimination law applicable in each jurisdiction.<sup>21</sup> There thus arises a research question with respect to disability

---

<sup>20</sup> An insightful discussion of intersectionality, noting the risk of over-simplifying its effects in responding to problems of injustice that result from machine learning technologies, appears in Hoffmann [25].

<sup>21</sup> The law concerning liability for disparate impact (often referred to outside the USA as indirect discrimination) has evolved differently between common law countries. See Khaitan [28] for a discussion.



discrimination in different legal and jurisdictional contexts regarding the implications of proxy discrimination, and what reforms, if any, should be introduced.<sup>22</sup>

At the core of the policy question raised by proxy discrimination is an ethical issue: under what circumstances is it morally permissible to discriminate against people based on a protected characteristic such as disability, if this characteristic is genuinely predictive of an outcome which is legitimately in the interests of the discriminator? This is a problem concerning the ethics of statistical discrimination [43]. Assuming that the costs of avoiding the discrimination by opting for a fairer but less predictively accurate AI solution are more than negligible, is the discriminator morally obligated to bear the costs and to choose the less discriminatory alternative? Depending on one's preferred normative analysis, the answer may be sensitive to details of the case at hand, for example whether the statistical relationships which purportedly justify the discrimination are in turn attributable to underlying social patterns of discriminatory practice [33].<sup>23</sup> In deciding what policy the law should reflect, and what choices should be made by developers and users of potentially discriminatory AI technologies, these moral issues are of central importance.

### 3.2.3 Human and Automated Decision-Making

Technical and policy measures that aim to reduce the discriminatory effect of a machine learning system are valuable, but they also presuppose a choice to develop the system for a specific purpose in the first place. This prior decision to use AI in a given context and the determination of what its role should be, if constructed, should also be examined in relation to the potential for discrimination against people with disabilities. It might on balance be preferable not to build the system at all, or to envision its role differently, thus shaping the character of the human–AI partnership.<sup>24</sup> Evidently, AI technology can be designed partly or completely to substitute for human judgment in making a specific kind of decision. What role, if any, the AI should have in a particular social situation is a choice that ought to be both well

---

<sup>22</sup> Prince and Schwarcz [41](§ IV.B) consider potential reforms, such as restricting the variables that may be used by AI systems in making certain kinds of decisions to a prescribed list of permitted factors.

<sup>23</sup> Selbst and Barocas [44](§ III.B) insightfully discuss difficulties resulting from the role of intuition in the reasoning required for the application of norms of non-discrimination. If the relations among variables apparently revealed by a machine learning system manifestly treat people with disabilities unfavorably, for example, but there is no coherent or plausible explanation of why this is the case, then evaluation of the grounds of these unequal outcomes becomes problematic. In some instances, techniques of 'interpretable' or 'explainable' machine learning may facilitate the emergence of a suitable explanation. On the other hand, and as the authors recognize, it would be naive to presuppose that social and natural phenomena are always amenable to explanations that cohere with human intuitions.

<sup>24</sup> An interesting further possibility is for a machine learning system to give an 'explanation' of its output that would enable an adversely affected person to change his or her situation sufficiently to achieve a more favorable classification. The difficulties of two promising approaches to such explanation are considered in Barocas, Selbst and Raghavan [4].

informed, and sensitive to the circumstances of the people involved as well as the rights and interests affected. It also raises issues that, if better understood, would allow for more effective policies and practices in deciding what part AI should play in different decision-making situations.

Competent and well-informed human decision-makers have capabilities of practical and moral reasoning that far surpass what is achievable by any AI technology yet devised. Human judgment can weigh and interpret the applicable ethical or legal norms, then apply them to the facts of a case to arrive at a just decision. The considerations taken into account need not be prescribed exhaustively in advance. General arguments have been developed in support of the view that, in the application of legal rules, each person has a moral claim for her or his case to be decided individually by the exercise of human judgment rather than to be determined algorithmically [6].<sup>25</sup> This position is supported by a number of independent philosophical arguments, for example regarding limitations on prior knowledge of uncertainties in the application of rules, the value of exercising discretion in decision making, and respect for each person's individuality [6](§ 3).<sup>26</sup>

An interesting further question suggested by this broader claim is whether a high degree of diversity present in a population, coupled with the need to make decisions based on disparate facts and norms that affect rights and interests, should be regarded as an additional ground for limiting the role of AI, even excluding it altogether, in reaching decisions. To develop the point more specifically, one may consider a hypothetical proposal to construct an AI system for determining, based on supplied data, whether specified support services requested by a person with a disability are likely to meet her or his needs. Such a system could be used either alone or, more probably, in combination with human review, by a government welfare program or in an educational setting. Arguably, the diverse nature of the population which would be subject to the proposed AI application, and the uniqueness of individual needs and circumstances, establishes a case for exercising human judgment that extends beyond the general arguments already cited. In a population that can reasonably be expected to contain many outliers, there is ample reason to be skeptical of efforts to formalize the decision-making problem and to develop algorithms capable of reaching just outcomes in most, let alone all cases. The diverse needs and circumstances of the people whose entitlements are to be determined establish a condition in which uncertainty in the interpretation and application of the relevant rules calls for modes of reasoning and consideration of unanticipated factors that only human decision making can provide. Justice may foreseeably require a degree of individual treatment of cases that current technology is unable to automate.

---

<sup>25</sup> Article 22 of the GDPR [17] establishes a limited right not to be subject to legally significant, fully automated decisions. For an argument against recognizing such a right to human involvement in individual decisions, which does not entirely address the philosophical grounds summarized in Binns [6], see Huq [26].

<sup>26</sup> Citron [11](§ III A and B) discusses the tendency of automation to substitute precise rules for more general legal standards that allow for the exercise of human discretion. This trend, Citron argues, prioritizes cost efficiency over justice.

Human judgment, however, is known to be fallible and prone to biases. Beyond intentional discrimination, which at least is under conscious control, prejudices against out-groups such as people with disabilities can be held and may influence behavior unconsciously [13]. Decision making can also be distorted by cognitive biases. If AI is combined with human involvement to reach decisions, automation bias [48] can limit the vigilance and effectiveness of the human decision-maker in identifying and compensating for erroneous findings of the AI technology. Invoking evidence from cognitive science and social psychology, Kleinberg et al. [29](§ 3) argue that human decisions are not only open to social and cognitive biases, but also that the true motivations are often not transparent to the agent.<sup>27</sup> Machine learning algorithms, they argue, can on the contrary be rigorously audited to ascertain the sources and extent of bias [29](§ 5). An appropriately designed algorithm can also be demonstrably less discriminatory than human judges, for example in criminal risk assessment [29](§ 6.2). They accordingly maintain that algorithmic decision making has distinctive equity-promoting advantages, noting [29](§ 6.1) the difficulty of determining the effectiveness of efforts to train humans to overcome biases.<sup>28</sup>

### 3.2.4 General Comments

Having regard to the unparalleled advantages of human judgment in making decisions in novel cases, and the potential of algorithms for auditability and bias mitigation, there is a need to develop a greater understanding of how best to gain the benefits of both in the service of social equality. Whereas technical measures can be taken to reduce biases in machine learning algorithms, social interventions can be made in an effort to overcome the more fundamental problem of human biases. The extent to which algorithmic bias can be detected and corrected in the face of a very diverse population of people with disabilities is an important question on which the potential of AI as a force for greater equality depends. If auditing is to be relied on as the principal mechanism, as proposed recently in Rambachan et al. [42](§ 1),<sup>29</sup> much depends on developing effective strategies for identifying and overcoming potentially context-specific manifestations of bias. Intersectional effects involving disability together with other protected characteristics, and the occurrence of proxy discrimination against possibly small subsets of the population, present two sources of difficulty. More generally, the many facets of diversity characteristic of disability constitute a challenge for overcoming the problem of algorithmic bias.

---

<sup>27</sup> A clear summary of the authors' position appears in Sunstein [50].

<sup>28</sup> A much discussed strategy for seeking to overcome human biases against social out-groups is the contact hypothesis. See, for example, Pettigrew and Tropp [39] and Pettigrew et al. [40].

<sup>29</sup> Under this proposal, the auditing is to be carried out by a regulator with the authority to compel changes that address discrimination. A more skeptical view of transparency as a means to greater accountability of machine learning systems is elaborated in Ananny and Crawford [2].

## 4 Conclusion

Developing AI technologies that facilitate equality while furthering the well-being and aspirations of people with disabilities is at least as much a social as it is a technical challenge. The concept of partnership between human beings and AI systems is useful in characterizing what ought to be built—a mutual interplay of social arrangements and software-based systems that promote morally good human ends through meeting practical needs. Devising meaningful approaches to including stakeholders who have disabilities (whether they be users or indirectly affected parties) in AI development projects is clearly a necessity. This participation should extend from the initial identification and clarification of the problem that is to be solved, through to the ultimate design, development, implementation, and maintenance of an AI-based solution. Further research and practical experience are vital to creating more specific guidance as to how this should be done, and as to what comprises an inclusive AI software development process. Promising approaches to design have emerged which empower potential users of the technology as well as indirect stakeholders whose interests are affected by its implementation, and which encourage reflection upon the value-dependent judgments associated with technical decisions. Such design-related scholarship offers valuable insights and methods, but treating the relationship of AI to people with disabilities as purely a design problem is not sufficient. Equally vital is the shaping of norms and policies associated with the development and use of AI.

The challenge of algorithmic bias raises technical, social, legal, and moral questions of importance in overcoming disability-based forms of discrimination that AI systems risk reinforcing. Many of these questions also apply to the problem of bias in machine learning generally, but there are disability-specific aspects and implications of these issues that have been emphasized here. Although technical means of mitigating or preventing bias have been proposed in recent literature, the perspective taken in this chapter suggests that any such measures should be applied in the context of a larger, policy-oriented approach to the problem. To a considerable degree, developing appropriate policy responses to issues of bias depends upon answering as yet unresolved research questions, some of which are identified in the preceding discussion.

Antidiscrimination law, at least in its predominant, adversarial and complaint-oriented form, seems inadequate, by itself, to redress harms resulting from algorithmic bias. There may thus need to emerge a complementary role, alongside antidiscrimination law, for proactive regulatory mechanisms which do not rely on people with disabilities who claim to have been adversely affected by algorithmic decisions to furnish the resources to sustain litigation. Proxy discrimination not only introduces practical difficulties for the removal or prevention of bias. It also raises moral and, depending on the applicable antidiscrimination regime, potentially also legal arguments purporting to justify discriminatory decision-making practices on the basis that an unbiased algorithm would be less effective in accomplishing a legitimate purpose, and, to that extent, more costly to the discriminator. The multiple social categories to which individuals belong may have unobvious intersectional consequences, involv-

ing disability together with other factors, which create additional sources of bias. Developing a deeper understanding of whether and to what extent this issue complicates efforts toward non-discrimination in the design and use of AI applications seems justified. Although acquiring data from people with disabilities is necessary to mitigate bias and to design AI applications that meet their needs effectively, it also risks compromising privacy and thus opens opportunities to exploit knowledge of an individual's disability. This is especially problematic for those who have limited capacity to give voluntary and informed consent to the use of their information, and under conditions in which the law allows data to be processed for the purpose of drawing inferences about a person's disability without the individual's knowledge or agreement.

The problem of bias also raises important research and policy questions concerning the appropriate roles to be accorded, respectively, to algorithmic and human decision making, particularly in application to the highly diverse circumstances of people with disabilities. Greater understanding is needed of how best to forge human-AI partnerships that overcome tendencies toward prejudice, biases, and discrimination as they manifest themselves in human decisions generally, as well as in the development and use of machine learning systems. Combining uniquely human capacities for practical reasoning and moral judgment appropriately with insights that can be derived from the operation of machine learning algorithms on large and diverse collections of data remains a challenge both in principle and in practice. An adequate response would proceed from an understanding of how biases occur in human judgment and in AI systems, while seeking to develop solutions that shape the social and policy-related aspects of the environments in which the technologies are developed and deployed, in addition to the technical design of the applications themselves.

Addressing these issues adequately in connection with disability can most effectively be pursued as part of a broader response to the potential for bias introduced by AI, and in particular by applications of machine learning. Devising appropriate human-AI partnerships should be regarded as a problem of putting in place effective policies, practices, technical expertise, and participatory processes throughout the development and maintenance of software projects. This is a large and complex undertaking, involving regulators, researchers, developers of AI technology, and, in the context at hand, people with disabilities.

**Acknowledgements** The author gratefully acknowledges Mark Hakkinen, Klaus Zechner, and Cary Supalo of Educational Testing Service for reviewing the manuscript. Mark Hakkinen and Kris Anne Kinney of Educational Testing Service offered valuable advice concerning creation of the diagrams. Anonymous reviewers contributed thoughtful suggestions for improving the chapter. This work has also been influenced by various seminars and workshops on the ethics of artificial intelligence that the author has attended.

## References

1. Alston, P.: Report of the special rapporteur on extreme poverty and human rights. Tech. rep., United Nations (2019). [https://www.ohchr.org/Documents/Issues/Poverty/A\\_74\\_48037\\_AdvanceUneditedVersion.docx](https://www.ohchr.org/Documents/Issues/Poverty/A_74_48037_AdvanceUneditedVersion.docx)
2. Ananny, M., Crawford, K.: Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* **20**(3), 973–989 (2018)
3. Barocas, S., Selbst, A.D.: Big data's disparate impact. *Calif. L. Rev.* **104**, 671 (2016)
4. Barocas, S., Selbst, A.D., Raghavan, M.: The hidden assumptions behind counterfactual explanations and principal reasons. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 80–89 (2020)
5. Berke, L., Caulfield, C., Huenerfauth, M.: Deaf and hard-of-hearing perspectives on imperfect automatic speech recognition for captioning one-on-one meetings. In: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 155–164 (2017)
6. Binns, R.: Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance* (2020). DOI <https://doi.org/10.1111/regg.12358>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/regg.12358>
7. Bogen, M., Rieke, A.: Help wanted: An examination of hiring algorithms, equity, and bias. Tech. rep., Upturn (2018). <https://www.upturn.org/reports/2018/hiring-algorithms/>
8. Bradshaw-Martin, H., Easton, C.: Autonomous or 'driverless' cars and disability: a legal and ethical analysis. *European Journal of Current Legal Issues* **20**(3) (2014)
9. Brewer, R.N., Kameswaran, V.: Understanding the power of control in autonomous vehicles for people with vision impairment. In: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 185–197 (2018)
10. Campbell, S.M., Stramondo, J.A.: The complicated relationship of disability and well-being. *Kennedy Institute of Ethics Journal* **27**(2), 151–184 (2017)
11. Citron, D.K.: Technological due process. *Wash. UL Rev.* **85**, 1249 (2007)
12. Crawford, K., Schultz, J.: Big data and due process: Toward a framework to redress predictive privacy harms. *BCL Rev.* **55**, 93 (2014)
13. Dasgupta, N.: Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social justice research* **17**(2), 143–169 (2004)
14. Degener, T.: A new human rights model of disability. In: V. Della Fina, R. Cera, G. Palmisano (eds.) *The United Nations convention on the rights of persons with disabilities*, pp. 41–59. Springer (2017)
15. Ehn, P.: Scandinavian design: On participation and skill. In: *Participatory design: Principles and practices*, p. 77. CRC Press (1993)
16. Employer Assistance and Resource Network on Disability Inclusion: Use of artificial intelligence to facilitate employment opportunities for people with disabilities (2019). [https://askearn.org/wp-content/uploads/2019/06/AI\\_PolicyBrief-A.pdf](https://askearn.org/wp-content/uploads/2019/06/AI_PolicyBrief-A.pdf)
17. European Union: Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). *Official Journal of the European Union* **L 119**, 1 (2016)
18. Findlater, L., Goodman, S., Zhao, Y., Azenkot, S., Hanley, M.: Fairness issues in ai systems that augment sensory abilities. *ACM SIGACCESS Accessibility and Computing* (125) (2020)
19. Friedman, B., Hendry, D.G.: *Value sensitive design: Shaping technology with moral imagination*. Mit Press (2019)
20. Friedman, B., Kahn, P.H., Borning, A.: Value sensitive design and information systems. In: K.E. Himma, H.T. Tavani (eds.) *The handbook of information and computer ethics*, pp. 69–101. John Wiley & Sons (2008)
21. Friedman, B., Nissenbaum, H.: Bias in computer systems. *ACM Transactions on Information Systems (TOIS)* **14**(3), 330–347 (1996)

22. Giermanowska, E., Raclaw, M., Szawarska, D.: *Employing People with Disabilities: Good Organisational Practices and Socio-cultural Conditions*, chap. 2, pp. 9–36. Springer (2020)
23. Guerreiro, J., Sato, D., Asakawa, S., Dong, H., Kitani, K.M., Asakawa, C.: Cabot: Designing and evaluating an autonomous navigation robot for blind people. In: *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 68–82 (2019)
24. Hoffman, S.: Big data's new discrimination threats: Amending the americans with disabilities act to cover discrimination based on data-driven predictions of future disease. In: I.G. Cohen, H.F. Lynch, E. Vayena, U. Gasser (eds.) *Big Data, Health Law, and Bioethics*. Cambridge University Press (2018)
25. Hoffmann, A.L.: Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* **22**(7), 900–915 (2019)
26. Huq, A.Z.: A right to a human decision. *Va. L. Rev.* **106**, 611 (2020)
27. Iglesias-Pérez, A., Loitsch, C., Kaklanis, N., Votis, K., Stiegler, A., Kalogirou, K., Serra-Autonell, G., Tzovaras, D., Weber, G.: Accessibility through preferences: context-aware recommender of settings. In: *International Conference on Universal Access in Human-Computer Interaction*, pp. 224–235. Springer (2014)
28. Khaitan, T.: Indirect discrimination. In: K. Lippert-Rasmussen (ed.) *The Routledge handbook of the ethics of discrimination*, pp. 30–41. Routledge (2017)
29. Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C.R.: Discrimination in the age of algorithms. *Journal of Legal Analysis* **10** (2018)
30. Kleiner, A., Kurzweil, R.C.: A description of the kurzweil reading machine and a status report on its testing and dissemination. *Bull Prosthet Res* **10**(27), 72–81 (1977)
31. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
32. Lewis, C.: Simplicity in cognitive assistive technology: a framework and agenda for research. *Universal Access in the Information Society* **5**(4), 351–361 (2007)
33. Lippert-Rasmussen, K.: Nothing personal: On statistical discrimination. *Journal of Political Philosophy* **15**(4), 385–403 (2007)
34. Loitsch, C., Weber, G., Kaklanis, N., Votis, K., Tzovaras, D.: A knowledge-based approach to user interface adaptation from preferences and for special needs. *User Modeling and User-Adapted Interaction* **27**(3-5), 445–491 (2017)
35. Marks, M.: Emergent medical data (2017). <https://blog.petrieflom.law.harvard.edu/2017/10/11/emergent-medical-data/>
36. Marks, M.: Algorithmic disability discrimination. In: I.G. Cohen, C. Shachar, A. Silvers, M.A. Stein (eds.) *Disability, Health, Law, and Bioethics*. Cambridge University Press (2020)
37. Marsch, N.: Artificial intelligence and the fundamental right to data protection: Opening the door for technological innovation and innovative protection. In: *Regulating Artificial Intelligence*, pp. 33–52. Springer (2020)
38. Nyholm, S., Smids, J.: Automated cars meet human drivers: Responsible human-robot coordination and the ethics of mixed traffic. *Ethics and Information Technology* pp. 1–10 (2018)
39. Pettigrew, T.F., Tropp, L.R.: A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology* **90**(5), 751 (2006)
40. Pettigrew, T.F., Tropp, L.R., Wagner, U., Christ, O.: Recent advances in intergroup contact theory. *International journal of intercultural relations* **35**(3), 271–280 (2011)
41. Prince, A.E., Schwarcz, D.: Proxy discrimination in the age of artificial intelligence and big data. *Iowa L. Rev.* **105**, 1257 (2019)
42. Rambachan, A., Kleinberg, J., Mullainathan, S., Ludwig, J.: An economic approach to regulating algorithms. Tech. Rep. w27111, National Bureau of Economic Research (2020)
43. Schauer, F.: Statistical (and non-statistical) discrimination. In: K. Lippert-Rasmussen (ed.) *The Routledge Handbook of the Ethics of Discrimination*, pp. 42–53. Routledge (2017)
44. Selbst, A.D., Barocas, S.: The intuitive appeal of explainable machines. *Fordham L. Rev.* **87**, 1085 (2018)
45. Shakespeare, T.: Critiquing the social model. In: A. Lawson (ed.) *Disability and equality law*, pp. 67–94. Routledge London (2017)

46. Shakespeare, T.: The social model of disability. In: L.J. Davis (ed.) *The disability studies reader*, 5 edn. Routledge (2017)
47. Shew, A.: Ableism, technoableism, and future ai. *IEEE Technology and Society Magazine* **39**(1), 40–85 (2020)
48. Skitka, L.J., Mosier, K., Burdick, M.D.: Accountability and automation bias. *International Journal of Human-Computer Studies* **52**(4), 701–717 (2000)
49. Spinuzzi, C.: The methodology of participatory design. *Technical communication* **52**(2), 163–174 (2005)
50. Sunstein, C.R.: Algorithms, correcting biases. *Social Research: An International Quarterly* **86**(2), 499–511 (2019)
51. Szarkowska, A., Krejtz, I., Klyszejko, Z., Wieczorek, A.: Verbatim, standard, or edited? reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American annals of the deaf* **156**(4), 363–378 (2011)
52. Treviranus, J.: The value of being different. In: *Proceedings of the 16th web for all 2019 conference—personalization-personalizing the web*, pp. 1–7 (2019)
53. Trewin, S.: Ai fairness for people with disabilities: Point of view. arXiv preprint [arXiv:1811.10670](https://arxiv.org/abs/1811.10670) (2018)
54. Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., Hebert, D., Lyckowski, N., Manser, E.: Considerations for ai fairness for people with disabilities. *AI Matters* **5**(3), 40–63 (2019)
55. UN: Convention on the rights of persons with disabilities. *United Nations Treaty Series* **2515**, 3 (2006)
56. United States: *K. w. v. Armstrong*, no. 14-35296 (9th cir. 2015)
57. Vanderheiden, G.C., Treviranus, J., Gemou, M., Bekiaris, E., Markus, K., Clark, C., Basman, A.: The evolving global public inclusive infrastructure (gpri). In: *International Conference on Universal Access in Human-Computer Interaction*, pp. 107–116. Springer (2013)
58. Whittaker, M., Alper, M., Bennett, C.L., Hendren, S., Kaziunas, L., Mills, M., Morris, M.R., Rankin, J., Rogers, E., Salas, M., et al.: Disability, bias, and ai. Tech. rep., AI Now Institute (2019). <https://ainowinstitute.org/disabilitybiasai-2019.pdf>
59. World Health Organization: Disability and health (2020). <https://www.who.int/news-room/fact-sheets/detail/disability-and-health>
60. World Institute on Disability: Ai and accessibility (2019). <https://wid.org/2019/06/12/ai-and-accessibility/>